

氏 名（本 籍）	山本 克彦（兵庫県）
学 位 の 種 類	博 士（工学）
学 位 授 与 番 号	甲 第 9 5 号
学 位 授 与 日 付	平成3 1 年3 月 2 5 日
専 攻	システム工学専攻
学 位 論 文 題 目	聴覚計算理論に基づく音声了解度客観評価指標
学位論文審査委員	（主査）教 授 入野 俊夫 （副査）教 授 和田 俊和 教 授 風間 一洋 荒木 章子（学外委員）

論文内容の要旨

1 研究の背景

雑音中に埋もれた音声の了解度を改善するために多くの雑音抑圧技術や音声強調技術が開発されているが、依然として実用的な客観評価指標は存在しない。そのため、被験者参加型の主観評価を行う必要があるが、時間や金銭面の問題から実験の実施規模に限界がある。音声了解度や音声品質の客観的評価指標の開発は、公共空間の音響設計や雑音抑圧処理手法の評価を効率的に行うためにも重要な課題である。

国際標準として推奨されている音声了解度指標（speech intelligibility index; SII）や音声伝達指標（speech transmission index; STI）といった音声了解度のための客観評価指標は、線形の伝送系における信号の劣化を仮定しているため、非線形な信号処理を含む雑音抑圧処理には対応できない。Jørgensen と Dau によって提案された音声振幅包絡パワースペクトルモデル（speech-based envelope power spectrum model; sEPSM）は、ガンマトーンフィルタバンクと変調フィルタバンクという2つの聴覚モデルから構成されており、スペクトルサブトラクションが適用された強調音声の了解度を予測できることが報告されている。しかし、音圧依存性を持たない線形なガンマトーンフィルタは難聴者をはじめとした個々人の聴覚特性を反映できない上に、sEPSM では“残留雑音”と呼ばれる雑音抑圧処理に依存した参照信号を推定する必要がある。

2 研究の目的と特徴

本研究の目的は、非線形な雑音抑圧処理を適用した強調音声に対する主観評価の結果を予測可能な、聴覚計算理論に基づく音声了解度の客観評価指標を開発することである。この目的を達成するために、sEPSM の問題点である (1) 線形なガンマトーンフィルタによる分析と (2) 雑音抑圧処理に依存した参照信号の問題を解決するための検討を行った。本研究の特徴は、人間の聴覚末梢系の非線形性（信号レベル依存の周波数選択性や圧縮特性）を表現することで、非線形な雑音抑圧処理手法を評価可能な音声了解度の客観評価指標を目指すことである。これにより、難聴者に限定した *ad-hoc* な枠組みを追加する必要が無く、健聴者と難聴者どちらの聴覚特性も模擬可能な評価指標になることが期待できる。また、従来の指標と同様のクリーン音声参照信号として使用するようアルゴリズムを再定義することによって、実用的な客観評価指標の開発を検討する。

3 研究の内容と結果

はじめに、非線形な動的圧縮型ガンマチャープ（dynamic compressive gammachirp; dcGC）フィルタバンクを用いて sEPSM を拡張した客観評価指標（dcGC-sEPSM）を提案した。dcGC フィルタバンクは、健聴者および難聴者のノッチ雑音マスキング実験の閾値データから、音圧依存性を含めた個々人の聴覚特性を直接反映することができる。dcGC-sEPSM は、dcGC フィルタバンクおよび変調フィルタバンクで分析された振幅包絡上の信号対雑音比（signal-to-noise ratio in the envelope domain; SNR_{env}）から音声了解度を予測する。ピンク雑音を重畳した雑音音声に、スペクトルサブトラクションとウィナーフィルタ型の雑音抑圧法（WF_{PSM}）を適用した強調音声の了解度を主観評価で測定し、客観的評価指標による予測結果との比較を行った。その結果、提案法の dcGC-sEPSM が WF_{PSM} 条件において、競合手法よりも優れた予測精度を示した。

次に、sEPSM における参照信号の問題を解決するために、振幅包絡上での信号対歪み比（signal-to-distortion ratio in the envelope domain; SDR_{env}）に基づくガンマチャープ振幅包絡歪み指標（gammachirp envelope distortion index; GEDI）を提案した（図 1(a)）。GEDI では、dcGC フィルタの出力として得られるクリーン音声と強調音声の振幅包絡間の歪み成分（図 1(b)）を算出し、その振幅包絡パワー（図 1(c)）を用

いて SDR_{env} を計算する。ピンク雑音を重畳した雑音音声と強調音声を用いて、GEDI による客観評価を行った。その結果、GEDI は dcGC-sEPSM と同様に WF_{PSM} 条件において、競手法よりも主観評価の結果を精度良く予測した。さらに、GEDI は sEPSM における参照信号の問題を解決することができた。

最後に、非定常雑音条件下で強調音声の了解度を予測可能な多重時間解像度版 GEDI (multi-resolution GEDI; mr-GEDI) を提案した。mr-GEDI では、変調フィルタバンクの出力を時間波形として抽出し、変調フィルタの中心変調周波数に依存した短時間フレームごとの振幅包絡パワーを計算することで SDR_{env} を算出する。提案法の評価を行うために、ピンク雑音条件下に加えてバブル雑音条件下の強調音声を使用して音声了解度の予測を行った。結果として、バブル雑音条件下およびピンク雑音条件下の両方において、mr-GEDI は主観評価の結果の傾向を正しく予測することができた。これにより、本論文で提案した客観評価指標の中で、mr-GEDI が最も実用性の高い客観評価指標であることが示された。

4 今後の課題と将来展望

本研究の成果により、聴覚モデル群の出力から算出される SDR_{env} と呼ばれる新たなアルゴリズムが音声了解度の客観評価に使用可能であることが示された。今後の課題として、音声了解度の予測精度を向上させるために、強調音声および参照信号の入力レベルの関係を適切に定義することや、聴覚計算理論に基づいた他の音響特徴量の検討が必要である。本研究で提案した客観評価指標を応用することで、将来的には難聴者を対象とした音声了解度予測、音声品質の客観評価指標への拡張、雑音抑圧処理手法のパラメータ最適化、補聴器をはじめとした聴覚補助機器のフィッティング最適化への貢献が期待できる。

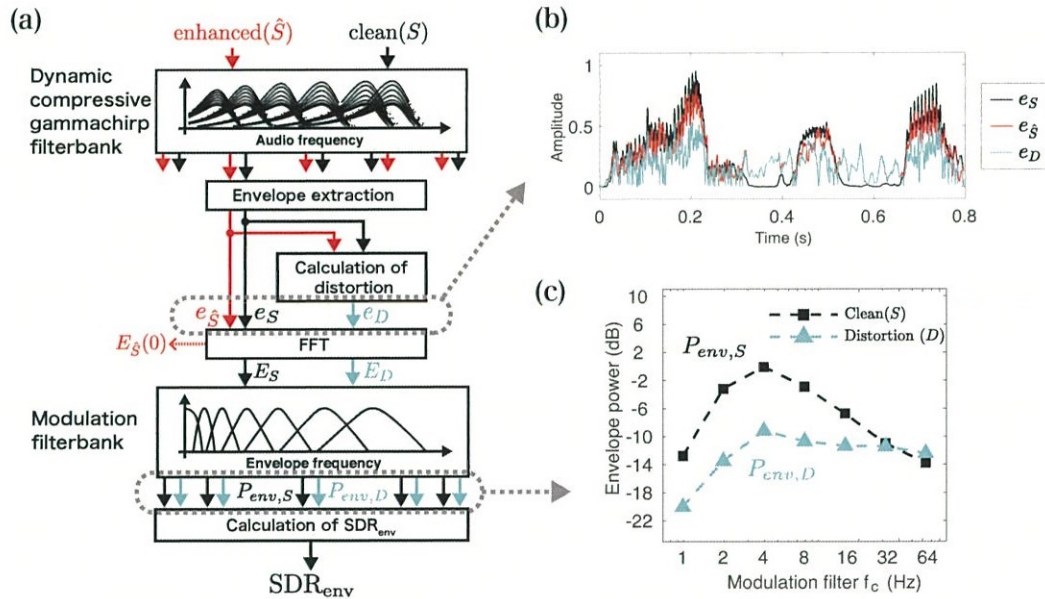


図1 本研究で提案した客観評価指標 GEDI の概要図. (a)GEDI の構成. (b) クリーン音声と強調音声の振幅包絡 (e_S と $e_{\hat{S}}$) と両信号から計算される振幅包絡歪み信号 (e_D) の例. (c) 変調フィルタバンクの出力として得られるクリーン音声と振幅包絡歪み信号の振幅包絡パワー ($P_{env,S}$ と $P_{env,D}$) の分布例.

論文審査の結果の要旨

本論文は、これまで客観的な評価が難しかった音声強調処理を施した音声の了解度に関して、新しい指標を提案している。人間の聴覚系のモデルを基礎とし、背景知識を説明したのち、1) 評価対象音声と残留雑音を用いた手法、2) 評価対象音声と雑音の無い原音声を用いた手法、3) これに多重解像度分析を加えて発展させた手法の提案をしている。評価実験は、音声強調処理手法（古典的、最先端の2種類）を用いて生成された強調音声を用いて、了解度予測の性能比較で行われた。この結果、これまで有力とされてきた数種類の従来手法に対し、本手法が優れていることが示された。このように新規性および有効性が高いことが客観的に示されている。また、予備審査等において問題点を指摘された箇所や、現状システムからの改善など今後の課題の記述に関する項目が、適切に加筆・修正されていた。これらのことより本論文は、博士論文に値すると判定した。

最終試験の結果の要旨

2月6日に開催された公聴会は、全審査員、学内関係者と数名の学外参加者が参加して行われた。発表に関しては問題なく、自らの研究内容を適確にプレゼンテーションした。さらに、周辺知識も含めた質疑応答も適切に行われ、十分な理解の上に研究を進めたことが審査員全員に理解できた。学位申請者は博士号を得るに足る学識・能力を有していると判断したため、最終試験は合格と判定した。